DOCUMENT RESUME

ED 370 376 FL 022 129

AUTHOR Russikoff, Karen A.
TITLE Hidden Expectations: Faculty Perceptions of SLA and ESL Writing Competence.
PUB DATE Mar 94
NOTE 20p.; Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (28th, Baltimore, MD, March 8-12, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; College Faculty; College Instruction; Comparative Analysis; *English (Second Language); Essays; Evaluation Criteria; *Grading; Graduation Requirements; Higher Education; Holistic Approach; *Interrater Reliability; Limited English Speaking; Native Speakers; State Universities; *Teacher Attitudes; Testing; *Testing Problems; *Writing Evaluation
IDENTIFIERS California

ABSTRACT

Problems inherent in the holistic scoring of essay examinations written by limited-English-speakers are examined, particularly in the context of one California state college in which English writing skills, holistically assessed, are required for graduation. These problems include lack of interrater reliability, raters' perceptions of their role, a reductive approach to scoring, imprecise criteria for scoring, confusion between inaccurate and non-standard structures, and clear prejudice based on the fact that the examinee was a student of English as a Second Language (ESL). A 1993 study of 392 faculty investigated teacher expectations of student writing in upper-and lower-division courses, including criteria for judging ESL student writing and beliefs about ESL student work in academic classes. Most respondents felt non-native speakers of English should meet the same criteria for English writing skills as native speakers, and declared that they graded ESL writers as they would native speakers. Implications are drawn in these areas: faculty awareness of different perceptions of writing proficiency; assessing student opportunities for learning; classroom teaching techniques; faculty cooperation on behalf of individual students; and both fairness and rigor in testing. Contains 25 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Karen A. Russikoff

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Hidden Expectations:

Faculty Perceptions of SLA and ESL Writing Competence

Karen A. Russikoff

ED 370 376

With the theme of the 1994 TESOL conference as "Sharing Our Stories," I am happy to be talking with you today since part of what I will be discussing involves the report of anecdotal evidence. Oftentimes this side of the story is dismissed since it is difficult to generalize and to quantify, but it is important since it conveys the human side of the issue, in this case, the effect of faculty on holistic scoring.

This type of student essay scoring has become commonplace in colleges and universities today for a variety of purposes, including placement, level movement, and exit from courses and programs, and while the the human element is an important factor for us always to remember, it is critical when the results are for high stakes assessment, such as in certifying competency for university graduation, which I will be discussing today.

Stock and Robinson (1987) remind us, "Testing, like teaching, is a social act with inevitable consequences" (p.119). For holistic assessment of ESL student writing, we need to consider these consequences carefully.

Let me begin this story-sharing with a tale that will provide some background. Once upon a time (1976, to be exact) in a kingdom by the sea (California, that is), a group of elders (known as the State University Board of Trustees) met and decreed that all students attending any of the now twenty state universities would be able to demonstrate writing proficiency in English before graduation. The decree became known as the GWAR, the Graduation Writing Assessment Requirement.

FL022129

BEST COPY AVAILABLE

...But then, a strange fog crept in across the land. The campuses could not agree on the form of assessment they should use. Consequently, each state university was allowed to create and administer its own writing instrument. Some used machine-scored objective tests; some provided course offerings as waivers or options. Most now use direct evaluation of student writing samples. A few allow 60 minutes; others up to four hours. On some campuses the writing is in response to a reading, and some request two separate essays to include different types of writing, for example, analysis, exposition or argumentation. All students take the test as they attain upper division, that is junior standing.

On one state university campus, the GWAR has become known as the Graduation Writing Test (The GWT). At the time of the test's inception on this campus in 1980, the exam included an objective test which was machine-scored and an essay which was holistically scored by faculty. However, after several years, it was determined that native speakers of English could not pass the objective grammar test (even though non-native speakers could) and so it was eliminated, leaving at present, only one 75-minute essay examination which is holistically scored.

Now a bit of background about this campus: The student population of this university reflects the changing demographics of the state of California. it is not especially unique. The recent campus census of the 17,500 students enrolled in fall 1993, cited a 63% minority population: i.e., approximately 32% Asian, 20% Hispanic, 6% Filipino, 4% African-American, and a small number of Native Americans. And while no records are kept on language backgrounds, it is apparent to faculty and administrators that the number of ESL students has significantly increased in recent years. Most of the nonnative-English-speaking students have citizenship or immigrant/refugee status, and only about 500 are visa students. The campus Test Office, which does request self-reported

information about student language backgrounds, cites the number of ESL students taking the GWT at any one sitting at approximately 50%.

This test is given once each quarter when between 1200 to 2500 tests are administered usually on a single Saturday morning. Students write in response to a prompt that has been field tested and is usually developed so that everyone will have the background knowledge to be able to respond, such as, discuss a piece of advice you might offer or were offered, followed by numerous examples, such as how to select a roommate, register for classes, care for a pet, paint a house, or start an exercise program. There is an effort to make the prompts bias-free, but as research suggests, there is much that we still do not know or fully recognize in order to fairly assist second language students, (e.g., Johns, 1991; Murphy & Ruth, 1993). The tests are scored two weeks following the test administration by 20 to 30 faculty members from the campus who spend an entire Saturday and Sunday involved in the holistic reading.

As you probably know, in the past holistic assessment of student essays resulted in a number of innovations, but today when we discuss this form of assessment, we generally refer to a process that combines criterion-referenced and norm-referenced scoring. The reader is supposed to focus on the writing in order to obtain a single overall impression that corresponds to a number on a scoring guide. Ed White, a key holistic advocate in the 1970's and 80's, confirmed holisticism (a term he coined) as judging the whole as being greater than the sum of its parts (1985). Yet the possibility of agreement among readers on what constitutes the appropriate "whole" has been a much disputed claim. Hamp-Lyons (1991) contended that the only agreement at present is that writing is "complex, multifaceted, and affected by cognitive and affective demands" (p.10).

In order to obtain a holistic score and reach the desired consensus, two steps are required: First, faculty are provided a scoring guide, the rubric, which categorizes types of writing. According to White, the purpose of this is to "set out standards for judgment so they can be explicit and debated" (1985: p. 7) Then, most often, the rubric is followed by the practice of scoring of papers that have previously been determined to be prototypical, that is they are calibrating essays or range finders. They represent aspects of the rubric not immediately evident without application.

The scoring guide for the test on this state university campus ranges from 1-6, with a 6 as a superior paper. The upper half descriptors are positive, general statements; the bottom half lists limitations and problems; all of these levels are then confirmed through the presentation and discussion of calibrating essays. The rubric, however, opens the door to another controversy of using faculty for holistic scoring. Winters (1980) argued that "No rubric can ever specify the entire set of criteria. There's always an 'X' factor which stands for how a reader interprets the rubric"(p. 78). And Janopoulos asserted that it is this vagueness, this uncertain aspect, that makes ESL teachers uncomfortable with holistic scoring and fearful that the evaluation of form will take precedence over communication (1993).

No distinctions between native and nonnative speakers of English are considered by the rubric used at this state campus, as both are supposedly assessed using these same criteria. Research has discussed concerns for sensitizing raters to ESL writing, but there is little agreement as to its effects (cf. Klammer, 1983; Breland & Jones, 1984, Ross, Burne, Callen, Eskey & McKay, 1984; Bochner, Albertini, Samar & Metz, 1992). Additionally, Hamp-Lyons (1991) warned that when L1 and L2 writing is mixed for the same holistic scoring, as it

is in this testing, that readers need further training to prevent them from concentrating on only low-level problems.

But what is it that readers do? How do we know? Because even in this advanced scientific age, it is still considered somewhat unethical to open a reader's head to see what is going on as he or she is reading, one of the few ways we really know what assessors are doing is through the discussion of the practice essays; this is when the rubric comes alive; this is the time when we find out what readers really notice and think about, what they expect and what they pounce on, as they silently read and rank student essays.

In a 1979 *New York Times* article, Edward Fiske said, "Tests don't judge people; people judge people." And nowhere is this more obvious than in holistic scoring.

Having participated in a number of holistic scoring sessions at the state university and for other purposes, this researcher has witnessed a wide range of reader judgments, depending upon the purpose of the test. In the state university's case, for meeting the graduation requirement, students must receive a minimum of a total score of 7 to pass (that is, two readers independently rate the paper from 1 to 6, and the scores are totalled.) Anyone receiving less than a 7 must take the test again in order to graduate. A waiver system is in place, but only a few students are willing to pursue the waiver since the result is a stamp on the student's transcript that says, "GWT Waiver Granted: Student did not pass the Graduation Writing Test." For a number of reasons that probably seem obvious to you, students are not eager to have this stigma permanently attached to their academic records. This is why we have ESL students who are taking the test for the 16th,17th, 18th time, sometimes years beyond their originally-scheduled graduation date. They return each quarter to take the test even when they are no longer taking courses.

Faculty all know the cut score for the test, so they are well aware that if they give a 3 to a paper that the student will probably not pass. The weight of such decisions might seem heavy because of such high stakes, that is whether or not a student graduates from the university, but faculty vary in their responses.

Although the holistic training in order to reach consensus is a rigorous 3 to 4-hour part of the first day, and is re-confirmed at intervals throughout the weekend, faculty adherence to the rubric is shiftable, and preconceived ideas seem to resurrect themselves throughout the session. In fact, some personal criteria are never completely contained. A number of researchers have even argued that it is probably unrealistic, unreasonable, and unfair to expect them to be. For example, Peter Elbow (1993) questions the brainwashing-versus-consensus issue of holistic evaluation; Barritt, Stock and Clark (1986) argue that we too easily dismiss discrepant scores when too much consensus, in the form of high inter-reader reliability, is what we really should be questioning.

During the holistic scoring, clusters of these shifting attitudes appear. Some holistic readers perceive their function as being the gatekeepers and guardians of the institution; that is they insist that <u>all students</u> must demonstrate equal minimal proficiency in order to reflect well upon the university when they are later assessed by employers. This attitude can be heard when a faculty member makes a comment such as,"If it can't be sent out to a client in its present form, then it will not pass with my score!" Nowhere on the rubric, nor in the range finder essays, is any such interpretation possible. Yet as this faculty member announces his bias aloud, others nod to confirm their own in light of his remark.

A more common reaction is when faculty readers narrow the focus to reductionistic concerns, with comments such as,

This paper is certainly ESL–I can tell by the handwriting;

Punctuation is merely an academic discipline;

It's too short.

All are disparaging and ultimately confirm a negative halo effect that influences their scoring. No one, however, can point to a single descriptor on the rubric that allows for this reductive approach.

Instead, the nearest application on this state university campus' rubric is the catch-all descriptor in #3 which says, "This score will be useful for papers that...are marred by more than a few minor grammatical inconsistencies." For most faculty, and in many instances, mechanical and grammatical are synonymous terms, and faculty readers often assume that if there are multiple forms of any type of writing mistake, the essay must fall into this broad category.

During follow-up discussions, the mechanics of an essay seem to attract a disproportionate amount of interest, yet discussions about grammatical control are much less likely than those about superficial problems of punctuation and spelling. One possible reason for this may lie in the fact that faculty are selected from a cross-section of academic disciplines; therefore, a professor from a discipline such as agriculture, biology, or electrical engineering might not be familiar with the vocabulary or the recognition necessary to indicate subject-verb agreement or the misuse of the present perfect tense or even article usage; what they see is something they perceive as non-standard English and what they interpret is that it is a mistake.

Basham and Kwachka (1991) confirmed this response to expectation as they found that when writing does not sound quite right to native speakers, they tend to assume it is wrong instead of different. Kaplan voiced the same concern over English native readers who may refuse "to interact with a text as the result of its 'foreignness,' " concluding that the nonnative student writer is thus "doomed to failure from the start."(1990, p. 15). Land and Whitely (1989) described this

condition as writing that is "out of focus" for native readers and they warn against such "rhetorical myopia" (p. 291).

A growing body of research, in fact, suggests that the general holistic impression that readers are supposedly seeking may not actually be the criteria on which they are basing their scores. On the contrary, they may subconsciously be counting surface errors, or relying on their own backgrounds, experiences, and biases. Vann, Meyer, Lorenz (1991) found that professional background accounted for tolerance levels for writing errors. According to their research, faculty in the so-called "hard sciences" (that is, physical, biological and mathematical) were less tolerant of language errors than were faculty from the "soft sciences"of humanities, education and social sciences. They hypothesized that it is the nature of these disciplines which is reflected in these attitudes; afterall, there is little room for growth or consideration of potential in the hard sciences, so none is attributed to language control.

This same study also noted that faculty who had the most exposure to ESL writing were the least critical of ESL errors (Vann et al., 1991). Hamp-Lyons confirmed the effect of experience, saying that ESL teachers have "the ability to recognize that even when content is at the mastery level, second language writers will still have language problems, sometimes even fossilized error patterns" (1991, p. 8).

Other comments reveal a general frustration and naivete by faculty when they say, "Why can't <u>these</u> students just spend a couple of months and learn how to write in English?" But when did a native student master this feat in only a couple of months? One faculty member recently announced at a scoring session that he recommended his students attend an intensive English language program for a while, and he concluded that the difference was amazing. But how long is "a while"? And in what context was it "amazing"? Who were these students and

what happened to them? These are not quantifiable nor clear answers to complex concerns. Yet his "quick fix" seemed to confirm other readers' attitudes, suspicions, and frustrations over ESL writing problems.

Readers' comments can offer considerable insight into more of the scoring process than one might guess. One specific scoring session culminated in a particularly startling realization because of the readers' comments. As usual, the reader training was managed with a number of calibrating essays to assist in understanding the correspondence of the readers' own strategies to the scores on the rubric. But the readers' discussion and their scoring quickly dwindled down to, "Oh! It's ESL...does that mean a 2 or a 3?" Puzzling over this reaction, this researcher began to count the immediately identifiable ESL essays that were offered as range finders. By the time the weekend was over, 11 ESL essays had been identified and offered as calibrators. Ten of the 11 were 3 and below scores (that is, they were representative of bottom half papers) and the 11th was considered a 3-4 split, right on the cusp. Thus, the faculty had been trained to expect a bottom-half score for any identifiable ESL paper, but the fact escaped everyone until the comments turned into immediate assignments of scores. After the reading, it was argued that the scores for ESL papers had been a full point lower than usual, but no changes were made.

In holistic readings for the GWT, faculty agree to forego their individual standards for the sake of reader consensus, that is reliable scoring. Yet during the session discussions of the range finders, and during breaks about the evaluation of "live" papers, the comments by faculty have seemed so far from the task at hand and so laden with multiple individual variables that this researcher suspected a comparable variation in their coursework assessment. Therefore, it became important to discover how faculty grade ESL students in their regular

academic courses and, along the way, to see if their beliefs supported their practices.

In the fall quarter of 1993, a questionnaire was sent to all faculty teaching on campus that quarter. A total of 392 surveys were completed and returned. The instrument was divided into five sections for analysis.

The first yielded information about faculty background, for example, years and levels of teaching, native languages and departments. There were few surprises in this portion of the returned surveys. Responses from all 47 departments in all six colleges and one school were received. Over half (53%)of the responding faculty had been teaching at the university for more than a decade and nearly half listed full professor as their rank; it may be surmised that the California State budget cuts have indeed been harsh to junior faculty. Further, 65 respondents were nonnative speakers and they listed 26 different languages as native.

The second section included questions that asked whether faculty were aware of ESL students in their classes. Ninety-four percent (n=369) responded that they have ESL students, implicitly stating that they understood the term. The few non-affirmative or nonresponsive answered by writing questions like, "What is an ESL student?"; "How would I know?"; "What does this mean?";and "Does the administration tell me what an ESL student is?" A cover letter had been provided by the Vice President of the University, and perhaps some attempt at political correctness may have been the reason for the reticence in answering the first question. All of these same respondents, however, answered Question 18 near the end of the survey that asked if ESL students encountered problems in their courses, to which they responded with either a yes or a no. Apparently, by the time they completed the survey they understood the term. Additional responses confirmed that 90% have ESL students every quarter, and for 60% of

the faculty between 11 and 50% of their classes are composed of second-language students. The overwhelming presence, number, and impact of ESL students were confirmed.

The third section asked about writing practices, that is whether faculty assign writing for coursework and what type, length and constraints affect it. To Question #4, 63% said they use it about the same as they did 2 to 3 years ago, before the budget cuts; 30% said more (attributable possibly to the recent writing across the curriculum project on campus;) and 6% said they use it less but blamed increased class size as the reason.

In Questions 7-12, a distinction was made between upper and lower division courses by pairing questions. Since the GWT is taken at the beginning of the junior year, soon after students complete usually only lower-division coursework, the fact that 30% do not use writing at all for this level is significant as it indicates fewer opportunities for writing practice and feedback. 40% never expect lengthy writing, nothing over 250 words, so even when students write they may be only completing fill-in-the-blank short answers. And 60% never expect students to write anything in class under a time limit, which is a serious factor on the 75-minute GWT, especially for second-language students.

On the other side, however, in upper division courses, students appear to be writing more. These responses indicated that 84% of the professors expect from 1 to more than 10 writing assignments per quarter (with the mode level at 3-6); 81% expect lengthy writing (over 250 words); and nearly half (48%) do expect students to write in class with a time limit.

The fourth section asked faculty for the criteria they use to judge ESL student writing. Not surprisingly (i.e., supporting research by Diedrich et al., 1961; Jacobs et al., 1981), content was the most important criterion for all faculty. Other

factors included organization, grammar, and mechanics, all at about the same degree of emphasis; and vocabulary and style were rated as least important.

The fifth section is perhaps the most interesting for our purposes today. This portion includes the questions seeking beliefs and feelings about ESL students and their work in academic classes. Question 13 went straight to the heart of the matter and asked: "In your opinion, should non-native speakers of English be required to meet the same criteria for English writing skills as native speakers of English?" Reponses were:

67%	Yes
17	No
12	Unsure
3	Other
1	No Response

In an attempt to confirm practice with this attitude, Question 14 sought to link the practice with the belief by asking, "How do you grade ESL student writing in comparison to native speakers of English?" The responses were:

29%	More leniently
0	More severely
64	Same
5	Other
2	No Response

What is particularly interesting about the responses to Questions 13 and 14 is not solely in the frequencies but in the fact that over half of those responding felt compelled to write something to clarify strong emotions, even when their answers

could have been answered with only a check box. The range of responses covered the extremes, demonstrated by responses such as these:

QUESTION 13

In your opinion, should non-native speakers of English be required to meet the same criteria for English writing skills as native speakers of English?

Definite:

"How could we expect less if they are to function in this society?"

"Yes, absolutely."

Qualified:

"The clumsiness of language associated with ESL is distinct from the clumsiness associated with other factors, e.g., lack of understanding, lack of study, etc."

"With compassion"

"ESL 'accent' okay."

Uncertain:

"It's probably unrealistic but we should strive for it."

"Not sure/uneven field of evaluation."

Question 14 offered similar categories of responses:

QUESTION 14

How do you grade ESL student writing in comparison to native speakers of English?

Definite:

"I do not grade students' writing skills."

"I don't grade anyone on their grammar."

"I always allow for some ESL differences."

Qualified:

"I look for content and ignore structural errors."

"I try to grade for intended meaning, not composition or spelling."

"I am more lenient on grammar only."

"Grammar, spelling, punctuation, vocabulary, forgiven for ESLs; I emphasize and grade content equally."

Uncertain:

"I correct all the spelling and grammar, but..."

"I give them many more chances to revise."

"I am not at all sure. If there is no attempt to fix the spelling and/or grammar mistakes, I get annoyed and may grade more severely. On the other hand, if mistakes are not so blatant, I tend to grade more leniently for those I assume to be ESL."

Questions 18 through 21 asked for faculty's awareness and understanding of their ESL students' academic progress. Question 18 asked if students encounter problems in their courses. Reponses were:

59%	YES
12	NO
27	UNSURE
2	NO RESPONSE

With a follow-up question, #19 asked what were the probable causes of any such problems. Results of checking all that apply included:

25%	Cultural problems
7	Inattention
3	Emotional Problems
37	Inadequate prior academic preparation
76	Language difficulties
4	Financial Concerns
2	Basic intelligence
7	Other

Question 20 asked for clarification to the assumed most common response in #19, that of language difficulties, and asked, "If language is a problem, what kinds of problems seem most significant?" Responses to check all that apply included:

68%	Weak writing skills
41	Understanding written questions
37	Responding orally to questions
36	Understanding spoken questions
36	Understanding lectures
35	Slow and inefficient reading
21	Weak communication with peers
17	Other
16	Slow/inexact note-taking

And then to discover what faculty already do to assist students and as a subtle suggestion for future options, Question 21 asked what they do to encourage ESL students to seek writing assistance. Responses were:

(n=55)	14%	a. I Don't
(201)	51	b. Suggest/require students get assistance
(211)	54	c. Suggest Learning Resource Center tutoring
(65)	16	d. Suggest EOP tutoring
(178)	45	e. Personally confer with/assist students myself
(70)	18	f. Match students with more capable peers
(55)	14	g. Suggest Reading Program tutoring in LRC
(66)	17	h. Suggest students read more for pleasure
(59)	15	i. Other

IMPLICATIONS:

While a "happily ever after" ending is usually sought at this point, at present this story does not appear to have one–maybe it should not. Perhaps what it really needs to be is an on-going saga. In that case, there are steps we need to take, things we need to do.

Stock and Robinson warn us, "It makes no sense to ask, 'How well do our students write?' unless we also ask, 'How well so we as assessors read?' " (1987; p. 119). The implications of this statement involve at least four areas for consideration:

1. For ourselves:

• We need to awaken to the awareness that we do not all agree about what constitutes writing proficiency, especially for L2 writers;

• We have to debunk the myth of faculty consensus of writing competence;

• We must find out if or where our students are not receiving all the opportunities for learning that we assume they are;

• We should understand that just because students have not revolted over this issue does not mean that it is acceptable. Does it always take a lawsuit or riot to change inequity?

2. For teaching:

• We need to recognize that the classroom is where students, especially ESL students who have fewer opportunities and less time to come to an understanding of academic writing, need to discover the criteria used for their writing assessment, not on the day of their graduation writing test--far too late in their academic careers!

• On certain campuses, faculty across the disciplines will have to drop the excuse, "We don't teach English" if they expect their students to demonstrate proficiency on a university-wide Graduation Writing Test–which is not the English Department's exit exam!

3. For sharing with our colleagues:

• We need to alert our peers to consider more than surface-level problems. This means we must make opportunities to present at in-service or faculty forums, to share knowledge and awareness about L2 students' learning requirements and strategies;

• We must make ourselves available to confer and assist other faculty's students when possible;

• We have to hear our peers' concerns and seek to develop some reasonable responses, such as adjunct courses.

4. For testing:

• Tests need to be regularly monitored and challenged. Changing demographics mean that tests normed on different populations, (for example, native speakers,) may not be appropriate for the present student population;

• We must be alert to the fact that what we assume a test is measuring may not be appropriate or the same for all test-takers;

• We are required to guard the test-takers and test-scorers as carefully as we guard the test standards.

According to Quellmalz (1980) at the UCLA Center for the Study of Evaluation, the original purpose of direct assessment of student essays was to create testing that would closely match performance objectives. We must ask

whether a graduation writing test that uses holistic assessment does indeed match what faculty expect in their coursework for ESL student writing. If not, we are left with one additional implication, which is that we must recognize that high-stakes evaluation means serious commitment to continual test evaluation (that is on-going, not every decade or two), including, and perhaps especially, focus on re-evaluation of the human effects.

References

Basham, C.S. & Kwachka, P.E. (1991). Reading the world differently: A cross-cultural approach to writing assessment. In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts (37-50). Norwood, NJ: Ablex.

Barritt, L., Stock, P., & Clark, F. (1986). Researching practice: Evaluating assessment essays. College Composition, 37, 315-327.

Bochner, J.H., Albertini, J.A., Samar, V.J., & Metz, D.E. (1992). External and diagnostic validity of the NTID Writing Test: An investigation using direct magnitude estimation and principal components analysis, Research in the Teaching of English, 26 (3),299-314.

Breland, H.N. & Jones, J.J.(1984). Perceptions of writing skills. Written Communication, 1 (1), 101-119.

Diedrich, P.B., French, J.W., & Carlton, S.T. (1961). Factors in judgments of writing ability. (Research Bulletin 61-15). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. 002-172).

Elbow, P. (1993). Ranking, evaluating, and liking: Sorting out three forms of judgment. College English, 55 (2), 187-206.

Fiske, E.B. (1979, November 18). Finding fault with the testers. New York Times Magazine, pp. 152-162.

Hamp-Lyons, L. (1991). Assessing Second Language Writing in Academic Contexts . Norwood, NJ: Ablex.

Huot, B.A. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. College Composition and Communication, 41 (2), 201-213.

Jacobs, H.L., Zinkgraf, S.A., Wormuth, D.R., Hartfiel, V.F., & Hughey, J.B. (1981). Testing ESL Composition: A Practical Approach. Rowley, MA: Newbury House Publisher, Inc.

Janopoulos, M. (1993). Comprehension, communicative competence, and construct validity: Holistic scoring from an ESL perspective. In M.M. Williamson and B. A. Huot (Eds.), Validating Holistic Scoring for Writing Assessment (pp. 303-325). Cresskill, NJ: Hampton Press, Inc.

Johns, A. (1990). Johns, A.M. (1991a). Interpreting an English competency exam: The frustrations of an ESL science student. Written Communication, 8 (3), 379-401.

Kaplan, R. (1990). Writing in a multilingual/multicultural context: What's contrastive about contrastive rhetoric. The Writing Instructor, 10, 7-18.

Klammer, T. (1983). Writing competencies: What should we expect from nonnative speakers? Long Beach, CA: English Council of the California State University.

Land, R.E. Jr., & Whitley, C. (1989). Evaluating second language essays in regular composition classess: Toward a pluralistic U.S. rhetoric. In D.M. Johnson and D.H. Roen (Eds.), Richness in Writing: Empowering ESL Students (pp.284-293). New York: Longman.

McKendy, T. (1992). Locally developed writing tests and the validity of holistic scoring. Research in the Teaching of English, 26 (2), 149-166.

Murphy, S. & Ruth, L. (1993). The field testing of writing prompts reconsidered. In M.M. Williamson and B. A. Huot (Eds.), Validating Holistic Scoring for Writing Assessment (pp. 266-302). Cresskill, NJ: Hampton Press, Inc.

Quellmalz, E. (1980). Problems in Stabilizing the Judgment Process. (CSE Report No. 136). Los Angeles: University of California.

Ross, S., Burne, K.G., Callen,J., Eskey, D., McKay, J. (1984). Expectations and evaluations of the second language student: Mattters of articulation in California education. A Report to the English Liaison Committee of the Articulation Conference of California.

Stock, P.L., & Robinson, J.L. (1987). Taking on testing: Teachers as tester-researchers. English Education, 19, 93-121.

Vann, R., Lorenz, F.O., & Meyer, D.M. (1992). Error gravity: Faculty response to errors in the written discourse of nonnative speakers of English. In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts (pp.181-196). Norwood, NJ: Ablex Publishing Corp.

Vaughan, C. (1992). Holistic assessment: What goes on in the reader's mind? In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts (pp. 111-126). Norwood, NJ: Ablex Publishing Corp.

White, E.M. (1985). Teaching and Assessing Writing. San Francisco: Jossey-Bass Publishers.

Williamson, M.M. & Huot, B. A. (Eds.) (1993). Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations. Cresskill, NJ: Hampton Press, Inc.

Winters, L. (1982). The effects of differing response criteria on the assessment of writing competence. Washington, D.C.: GPO. (ERIC Document Reproduction Service No. ED 212 659).